SONDERHEFTE ZUM ALLGEMEINEN STATISTISCHEN ARCHIV
Organ der Deutschen Statistischen Gesellschaft
Heft 21

Recent Trends in Statistics

Proceedings of the
Anglo-German Statistical Meeting
Dortmund 24-26 May 1982

Edited by Siegfried Heiler

With Contributions of

R. Bailey, J. A. John, F. Pukelsheim, H. Stenger,
R. A. Sugden, P. Robinson and H. J. Lenz, G. Rothe,
N. Schmitz, P. Diggle, V, Isham, W. Fieger, M. Deistler,
G. Tunnicliffe-Wilson, A. J. Lawrance, E. J. Hannan

VANDENHOECK & RUPRECHT IN GÖTTINGEN

Sonderhefte zum Allgemeinen Statistischen Archiv

Heft 1: Neuere Entwicklungen auf dem Gebiet der Zeitreihenanalyse, 1970
Heft 2: Zur Reform der amtlichen Industriestatistik, 1971
Heft 3: Zur Messung von Erträgen, 1973
Heft 4: Neuere Methoden der Produktivitätsmessung, 1973
Heft 5: Aufstellung und Analyse von Input-Output-Tabellen, 1973
Heft 6: Stand der Einkommensstatistik, Individual- und Haushaltseinkommen, Einkommensschichtung, 1974
Heft 7: Konjunktur-Indikatoren, 1975
Heft 8: Der Baumarkt im Wandel und seine statistische Messung, 1975
Heft 9: Beiträge zur Zeitreihenanalyse,1976
Heft 10: Messung der Kaufkraft des Geldes, 1976
Heft 11: Statistiken der Erwerbstätigkeit und Beschäftigung, 1977
Heft 12: Makroökonometrische Modelle für die Bundesrepublik Deutschland, 1978
Heft 13: Grundlagen und Methoden der Verbrauchsstatistiken und der Konsumforschung 1978
Heft 14: Splinefunktionen in der Statistik, 1978
Heft 15: Statistische Informationssyteme, 1979
Heft 16: Methodik und Anwendung ökonometrischer Entscheidungsmodelle, 1980
Heft 17: Praktische Anwendungen von Stichprobenverfahren, 1980
Heft 18: Die Weiterentwicklung der Input-Output-Rechnung in der Bundesrepublik Deutschland, 1982
Heft 19: International Use of Input-Output Analysis, 1982
Heft 20: Qualitative und quantitative Konjunkturindikatoren, 1983

CIP-Kurztitelaufnahme der Deutschen Bibliothek

Recent trends in statistics : proceedings of the Anglo-German Statist. Meeting, Dortmund, 24–26 May 1982 / ed. by Siegfried Heiler. With contributions of R. Bailey ... – Göttingen : Vandenhoeck und Ruprecht, 1983.
(Sonderheft zum Allgemeinen statistischen Archiv ; H. 21)

ISBN 3-525-11193-2

NE: Heiler, Siegfried [Hrsg.]; Bailey, Rosemary [Mitverf.]; Anglo-German Statistical Meeting ⟨01, 1982, Dortmund⟩; Allgemeines statistisches Archiv / Sonderhefte

 Druck: Hubert & Co., Göttingen

Optimal Designs for Linear Regression

By FRIEDRICH PUKELSHEIM, Freiburg im Breisgau

Abstract. A brief survey is given of characterizations of optimal experimental designs in the approximate design theory. These may be useful also for the exact design theory, as is demonstrated with some new results for two-way block designs.

1. Introduction.

In experimental design theory a design ξ is taken to be a probability measure with finite support on a design space $\mathfrak{X}$, with the interpretation that a proportion $\xi(x)$ of all observations is to be drawn under experimental conditions $x \in \mathfrak{X}$. In the exact theory, the weights $\xi(x)$ are restricted to be of the form 0, 1/n, ..., (n-1)/n, 1, and this leads to designs which are realizable for sample sizes n, 2n, etc. In the approximate theory, $\xi(x)$ may attain any value between 0 and 1 and thus, in general, only provides an approximation to a design which is realizable. But many problems are discussed more easily in the approximate, rather than the exact theory.

Traditionally, the examples to the theory are subdivided into two main classes, one where the design space $\mathfrak{X}$ is discrete, and one where $\mathfrak{X}$ is a continuum. However, the discreteness of a design space, and the discreteness inherent in the exact theory approach are of a quite different nature. Therefore both, a discrete and a continuous design space may be analyzed in either the exact or the approximate theory, as Section 3 illustrates by example. In Section 2 we first give an appropriate characterization of optimal designs for linear regression.

2. Optimality characterization.

The objective of optimal design theory is to find designs ξ which provide a maximum of information on the unknown model parameters. We shall now specify the assumptions which allow to make this objective more precise.

Suppose that under every experimental condition x in the design space $\mathfrak{X}$ we can make a real observation Y_x which follows a normal distribution with mean $f(x)'\beta$ and variance σ^2. Here the regression function f on $\mathfrak{X}$ is assumed known, taking values in $\mathbb{R}^k$ such that its image $f(\mathfrak{X})$ is compact, while $\beta \in \mathbb{R}^k$ and $\sigma^2 > 0$ form the unknown parameters. All observations are independent, regardless of whether they are drawn under identical or under distinct experimental conditions.

Standard linear model considerations, as detailed in Krafft (1978) or Silvey (1980), justify to compare two designs ξ and η by means of their information matrices $M(\xi)$ and $M(\eta)$, defined by

$$M(\xi) = \int_{\mathfrak{X}} f(x)f(x)'d\xi \; .$$

More generally, when K is a prescribed k by s matrix of full column rank s, the information matrix for $K'\beta$ is given by

$$C(M) = (K'M^{-}K)^{-1}$$

provided $K'\beta$ is identifiable (estimable, testable) under M. A formal criterion for identifiability is that the range of the matrix M contains the range of K, and in this case the s by s matrix C(M) is well defined and positive definite; otherwise we simply set C(M) = 0.

Only in rare cases will it be possible to maximize information matrices in the partial ordering of non-negative definite matrices. In general, we must further specify a real optimality criterion j(C) in order to make the problem amenable to a solution. We shall take j to be one of the functions

$$j_p(C) = (\text{trace } C^p/s)^{1/p} , \qquad j_0(C) = (\det C)^{1/s} ,$$

i.e., $j_p(C)$ is the generalized mean of order $p \in (-\infty,1]$ of the (positive) eigenvalues of C.

Now let $\mathfrak{M}$ be a convex compact set of non-negative definite k by k matrices, a subset of all information matrices. The <u>optimal design problem</u> then reads:

$$\text{Maximize } j_p(C(M)), \text{ subject to } M \in \mathfrak{M}.$$

A matrix $M \in \mathfrak{M}$ which solves this problem will be said to have $\mathfrak{M}$-maximal j_p-information for $K'\beta$. Notice that this maximization problem is in terms of, not designs ξ, but information matrices M. And it is a problem of maximizing information, rather than one of minimizing risk.

Theorem 1. Let $M \in \mathfrak{M}$ be an information matrix under which $K'\beta$ is identifiable. Then M has $\mathfrak{M}$-maximal j_p-information for $K'\beta$ if and only if there exists some k by k matrix G with $MGM = M$ such that, with $B = G'KC^{p+1}K'G$ and $C = C(M)$,

$$\text{trace } AB \leqq \text{trace } C^p , \qquad \text{for all } A \in \mathfrak{M} .$$

The key feature of this optimality characterization is that the competing information matrices A enter it linearly, and that inversions are required of the optimality candidate M, only.

When M is non-singular, then M^{-1} is the unique choice for G and Theorem 1 may be derived by differential calculus (Kiefer 1974, p. 865). The general proof is based on duality theory of convex analysis (Pukelsheim 1980, p. 356); the generalized inverse G of M which appears in the characterization is constructed via the dual of the optimal design problem. Alternative proofs use subdifferential calculus, the Strong Lagrangian Principle, or directional derivatives (Pukelsheim & Titterington 1982).

3. Two-way block designs.

We now turn to an application of the approximate theory to a discrete design space. Suppose the experimental conditions are reflected by a triplet (i,j,k), where i is one of v varieties, j is one of b blocks of a first factor, and k is one of c blocks of a second factor. The design space then is

$$\mathfrak{X} = \{1,\ldots,v\} \times \{1,\ldots,b\} \times \{1,\ldots,c\} \ ,$$

and a design ξ may be taken to be a stochastic vector of vbc dimensions, with entries $\xi(i,j,k)$ in lexicographic order.

In the model for the elimination of two-way heterogeneity, the variety effects α_i, and the block effects γ_j and δ_k are assumed non-random and additive:

$$Y_{i,j,k,l} = \alpha_i + \gamma_j + \delta_k + \varepsilon_{i,j,k,l} \ .$$

The error terms are independent and identically distributed according to a normal distribution with mean 0. The vector parameter β and the regression function f thus are

$$\beta = \begin{pmatrix} \alpha \\ \gamma \\ \delta \end{pmatrix} \in \mathbb{R}^{v+b+c} \ , \qquad f(i,j,k) = \begin{pmatrix} e_i \\ e_j \\ e_k \end{pmatrix} ,$$

where e_i etc. is the i-th Euclidean unit vector of appropriate order.

Given a design ξ, let $r \in \mathbb{R}^v$ be the variety marginals, i.e., $r_i = \Sigma_j \Sigma_k \xi(i,j,k)$, and let $s \in \mathbb{R}^b$ and $t \in \mathbb{R}^c$ be the block marginals of the first and the second factor, respectively. Further denote by W_1, W_2, and W_{12} the two-dimensional marginals of varieties with blocks of the first factor, of varieties with blocks of the second factor, and of blocks of the first with blocks of the second factor. Writing Δ_r for a diagonal matrix with r on the diagonal, the information matrix M of ξ then is completely determined by the two-dimensional marginals, according to

$$M = \begin{bmatrix} \Delta_r & W_1 & W_2 \\ W_1' & \Delta_s & W_{12} \\ W_2' & W_{12}' & \Delta_t \end{bmatrix} = \begin{bmatrix} \Delta_r & W \\ W' & E \end{bmatrix}, \text{ say.}$$

As usual, the parameters of interest are the <u>variety contrasts</u> $(\alpha_1-\bar{\alpha}.,\ \ldots,\ \alpha_v-\bar{\alpha}.)' = [I_v-J_v/v \;\vdots\; 0]\beta$, with J_v the v by v matrix each entry of which is unity. Denote by A^+ the Moore-Penrose inverse of the matrix A.

Lemma 1. Let M be an information matrix under which the variety contrasts are identifiable, and choose a symmetric generalized inverse $\bar{F}$ of $F = \Delta_t-W_{12}'\Delta_s^+W_{12}$. Then the information matrix $C = C(M)$ for the variety contrasts is

$$C = \Delta_r - W_1\Delta_s^+W_1' - (W_2-W_1\Delta_s^+W_{12})\bar{F}(W_2-W_1\Delta_s^+W_{12})' .$$

Thus C may be read as a Schur complement of F in $\begin{bmatrix} H & G \\ G' & F \end{bmatrix}$, where F as well as $G = W_2-W_1\Delta_s^+W_{12}$ and $H = \Delta_r-W_1\Delta_s^+W_1'$ themselves are Schur complements. The proof of Lemma 1 parallels the argument in Krafft (1978, p. 220) who considers a special case. For block-block product designs, i.e., when $W_{12} = st'$, the matrix C simplifies to $C = \Delta_r - W_1\Delta_s^+W_1' - W_2\Delta_t^+W_2' + rr'$, since $F = \Delta_t-tt'$ allows the choice $\bar{F} = \Delta_t^+$.

It is now easy to establish optimality of <u>variety-block product designs</u>, by definition these designs have $W_1 = rs'$ and $W_2 = rt'$. A design with uniform variety marginals will be said to be <u>equi-replicated</u>. Denote by Ξ the set of all two-way block designs ξ, and by $\Xi(r_o)$ its subclass with positive variety marginals r_o given. Our next result also follows from the general theory (cf., Pukelsheim 1982, Thm. 4), but we here give a simpler direct proof. Recall that uniform optimality requires an information matrix which is maximal in the partial matrix ordering, while universal optimality (Kiefer 1975, p. 334) means maximal trace as well as C being positively proportional to $K_v = I_v-J_v/v$.

Theorem 2. (a) Variety-block product designs with variety marginals r_o are the only uniformly optimal designs in $\Xi(r_o)$ for the variety contrasts, their common C-matrix is $\Delta_{r_o} - r_o r_o'$. (b) Equi-replicated variety-block product designs are the only designs which are universally optimal in Ξ for the variety contrasts, with $C = K_v/v$.

Proof. (a) The product design $\xi = r_o \otimes s \otimes t$ has C-matrix $C(\xi) = \Delta_{r_o} - r_o r_o'$. For every other design η with variety marginals r_o we may choose $\bar{F}$ in Lemma 1 to be non-negative definite. Then

$$C(\eta) \leqq \Delta_{r_o} - W_1 \Delta_s^+ W_1' = \Delta_{r_o} - r_o r_o' - (W_1 - r_o s')\Delta_s^+(W_1 - r_o s')'$$

$$\leqq \Delta_{r_o} - r_o r_o' = C(\xi)\ .$$

Thus ξ is optimal. Equality holds in the second inequality if and only if $W_1 = r_o s'$ and, by symmetry, $W_2 = r_o t'$. In this case $W_2 - W_1 \Delta_s^+ W_{12} = 0$, and this forces equality in the first inequality. (b) The Cauchy Inequality immediately gives $\mathrm{trace}(\Delta_{r_o}' - r_o r_o') = 1 - r_o' r_o \leqq 1 - 1/v$.

As an example, consider the variety contrasts for 4 varieties in 6 by 6 blocks. Krafft (1978, p. 371f.) quotes a generalized Youden design GYD, exact for sample size 36, with determinant information $j_o(\mathrm{GYD}) = 0.23148$, and another exact design OED* with information $j_o(\mathrm{OED}) = 0.23175$. By Theorem 2(b), an optimal approximate design is $\mathrm{OAD} = 1_4 \otimes \begin{bmatrix} 1_3 \\ 0_3 \end{bmatrix} \otimes \begin{bmatrix} 1_3 \\ 0_3 \end{bmatrix} /36$, with information $j_o(\mathrm{OAD}) = 0.25 = 1/v$. And OAD is exact for 36 observations!

Variety-block block-block product designs are j_p-optimal even for a certain <u>maximal</u> system of parameters. The proof of this makes use of Theorem 1 and parallels that for one-way block designs (cf., Pukelsheim 1982, Thm. 5). This parallelism breaks down as soon as incomplete designs are considered.

---

*Recently shown to be optimal for size 36 among all designs with uniform block-block marginals. (Personal communication from O. Krafft)

Indeed, suppose GYD is a generalized Youden design and let $S_1 \subset \mathfrak{X}_1 = \{1,..,v\} \times \{1,..,b\}$ and $S_2 \subset \mathfrak{X}_2 = \{1,..,v\} \times \{1,..,c\}$ be the sets of points which support the variety-block marginals W_1 and W_2, respectively. Denote by $\Xi(S_1,S_2)$ the set of all two-way designs whose variety-block marginals have a support contained in S_1 and S_2, respectively. As far as variety contrasts are concerned, GYD may fail to be optimal in $\Xi(S_1,S_2)$. This is a marked difference to the behaviour of balanced incomplete block designs in one-way models (cf., Pukelsheim 1982, Thm. 7).

Non-optimality of generalized Youden designs. We choose a model for 3 varieties in 6 by 6 blocks. Let $D(\xi)$ be a 6 by 6 array whose (j,k)-entry is the set of varieties to be observed with block-block combination (j,k). Consider the designs:

$$D(GYD) = \begin{matrix} 3 & 3 & 2 & 2 & 3 & 2 \\ 3 & 1 & 1 & 3 & 3 & 1 \\ 2 & 1 & 2 & 2 & 1 & 1 \\ 2 & 3 & 2 & 3 & 3 & 2 \\ 3 & 3 & 1 & 3 & 1 & 1 \\ 2 & 1 & 1 & 2 & 1 & 2 \end{matrix} \qquad D(BAD) = \begin{matrix} 3+2 & 3 & 2 & 2+3 & 3 & 2 \\ 3 & 1+3 & 1 & 3 & 3+1 & 1 \\ 2 & 1 & 2+1 & 2 & 1 & 1+2 \\ 2+3 & 3 & 2 & 3+2 & 3 & 2 \\ 3 & 3+1 & 1 & 3 & 1+3 & 1 \\ 2 & 1 & 1+2 & 2 & 1 & 2+1 \end{matrix}$$

$$D(OAD) = \begin{matrix} 3+2 & \emptyset & \emptyset & 2+3 & \emptyset & \emptyset \\ \emptyset & 1+3 & \emptyset & \emptyset & 3+1 & \emptyset \\ \emptyset & \emptyset & 2+1 & \emptyset & \emptyset & 1+2 \\ 2+3 & \emptyset & \emptyset & 3+2 & \emptyset & \emptyset \\ \emptyset & 3+1 & \emptyset & \emptyset & 1+3 & \emptyset \\ \emptyset & \emptyset & 1+2 & \emptyset & \emptyset & 2+1 \end{matrix}$$

GYD is a generalized Youden design, BAD is a better and OAD is an optimal design; their C-matrices are of the form ρK_3 with

$$\rho(GYD) = 1/6 < \rho(BAD) = 1/5 < \rho(OAD) = 1/4 .$$

Theorem 1 may be used to prove that OAD is universally optimal for the variety contrasts in $\Xi(\mathfrak{X}_1,S_2)$, and in $\Xi(S_1,\mathfrak{X}_2)$. The above designs are exact for size 36, 48, and 24; and the ranks of their information matrices are 13, 13, and 11; respectively.

References.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). **The Annals of Statistics 2**, 849-879.

Kiefer, J. (1975). Construction and optimality of generalized Youden designs. In: J. N. Srivastava, (ed.), **A Survey of Statistical Design and Linear Models.** Amsterdam, North-Holland, 333-353.

Krafft, O. (1978). **Lineare statistische Modelle und optimale Versuchspläne.** Göttingen, Vandenhoeck & Ruprecht.

Pukelsheim, F. (1980). On linear regression designs which maximize information. **Journal of Statistical Planning and Inference 4**, 339-364.

Pukelsheim, F. (1982). On optimality properties of simple block designs in the approximate design theory. **Journal of Statistical Planning and Inference**, forthcoming.

Pukelsheim, F. & Titterington, D.M. (1982). General differential and Lagrangian theory for optimal experimental design. In preparation.

Silvey, S.D. (1980). **Optimal Design.** London, Chapman and Hall.

F. Pukelsheim
Institut für Mathematische Stochastik
Albert-Ludwigs-Universität
Hebelstraße 27
D-7800 Freiburg im Breisgau
Federal Republic of Germany